



Evolution of a Common Structural Core in the Internal Ribosome Entry Sites of Picornavirus

SHU-YUN LE & JACOB V. MAIZEL, JR.

Laboratory of Mathematical Biology, Division of Cancer Biology Diagnosis and Centers, National Cancer Institute, NIH, Bldg 469, Room 151, Frederick, Maryland 21702
E-mail: shuyun@fcrlv1.ncifcrf.gov

Abstract. The translational control involving internal ribosome binding occurs in poliovirus (PV), human rhinoviruses (HRV), encephalomyocarditis virus (EMCV), foot-and-mouth disease virus (FMDV), and hepatitis A virus (HAV). Internal ribosome binding utilizes *cis*-acting genetic elements of approximately 450 nucleotides (nt) termed the internal ribosome entry sites (IRES) found in these picornaviral 5'-untranslated region (5'UTR). Although these IRES elements are quite different in their primary sequence, a similar folding structure with a conserved 3' structural core exists in the IRES. Phylogenetic analysis and RNA folding of the 5' UTR of picornaviruses, including PV types 1–3, coxsackievirus types A and B, swine vesicular disease virus, echoviruses, enteroviruses (human and bovine), HRV, HAV, EMCV, mengovirus, Theiler's murine encephalomyelitis viruses, FMDV, and equine rhinoviruses, indicates that the predicted conserved structural core is indeed a general structural feature for all members of the picornavirus family. The evolution of a common structural core likely occurred by the gradual addition or deletion of structural domains and elements to preserve a similar tertiary structure that facilitates the utilization of the IRES in specific host-cell environments.

Key words: internal ribosome entry site, initiation of translation, picornaviridae, phylogeny, evolution

Introduction

The Picornaviridae comprise one of the largest and most important families of human and agricultural pathogens. Foot-and-mouth disease virus, a member of this family is one of the most important single pathogens of livestock. In general, picornaviruses are closely related and can be divided into five genera: (a) enteroviruses that includes human poliovirus types 1–3 (PV 1–3), human coxsackievirus types A and B (Cox-A and Cox-B), swine vesicular disease virus (SVDV), human echoviruses (Echo), human enteroviruses (HETV), and bovine enteroviruses (BEV), (b) rhinovirus, human rhinoviruses (HRV), (c) cardiovirus, including encephalomyocarditis viruses (EMCV), mengovirus (Mengo) and Theiler's murine encephalomyelitis viruses (TMEV), (d) aphthovirus, foot-and-mouth disease viruses (FMDV), (e) hepatovirus, human hepatitis virus A (HAV). Members of the picornavirus family, equine rhinoviruses 1 and 2 (ERV-1 and ERV-2) have not yet been assigned genera. All have a relatively long 5' untranslated region (5'UTR), from 600 to more than 1060

nucleotides (nt), and multiple, unused AUG triplets. Evidence from numbers of laboratories indicates that the translation of PV (1–2), EMCV (3–4), TMEV (5), HRV (6), FMDV (7) and HAV (8–9) is initiated by a mechanism of ribosome-binding and cap-independent translation initiation (10,11). The mechanism of direct internal ribosomal entry depends on a *cis*-acting element termed the IRES that extends over approximately 450 nt within the viral 5'UTR. The IRES elements in PV and EMCV are located within the boundaries of nt ~130 to ~600, and nt ~400 to ~830, respectively. Although both IRES elements perform the same function for each virus, little sequence similarity was observed between them. These picornavirus IRES elements are highly structured. In the previous paper (12) we have proposed complex RNA secondary structural forms in the IRES elements of PV, Cox-A and Cox-B, HRV, EMCV, TMEV, FMDV and HAV sharing a similar folding shape. Especially prominent is a common RNA structural core with similar three-dimensional forms in the 3' portion of these IRES elements. The IRES core element extends 140 to 170 nt and is

dependent on their genera. The phenomenon is of general importance because the main structural features in the structural core of picornavirus IRES is also conserved in the IRES elements of hepatitis C virus (HCV) and pestiviruses consistent with the fact that these viral IRES are functionally related to each other. As a result, the common structural core was suggested to play a crucial role in the mechanism of direct internal ribosome entry (12).

In this article, we focus on the analysis of genetic relation of the IRES elements between picornaviral genera. We perform phylogenetic analysis for the 5'UTR of all picornaviral RNAs found in the current GenBank and EMBL databases. The proposed models of the highly conserved structural core for these IRES elements are examined in terms of their evolutionary relationships.

Materials and Methods

The analyzed 5'UTR picornavirus sequences are listed in Table 1. To determine the IRES elements in those picornaviruses of Echo, HETV, BEV, SVDV, Mengo, ERV and other members of picornaviruses both sequence comparison and RNA folding were performed. In the sequence comparison the NUCALN (13) and MAL (14) programs were employed. The RNA folding of 5'UTR sequence and predicted IRES core elements were searched using the EFFOLD program (15). The computed thermodynamically favored base-pairing regions were examined by phylogenetic comparative analysis.

In the phylogenetic analysis, the alignment of homologous nucleotides in the same subgenera was first completed using program MAL. The alignment of nucleotides in each genera was refined by editing the previous alignment data and by using the common folding information (12). In the evolutionary analysis of these picornaviral IRES core elements, evolution is conventionally described in terms of a unrooted bifurcating tree that was determined by the DNAPARS and CONSENSE tree programs in the PHYLP 3.5C package (15).

Results and Discussion

Structural Conservation of Picornaviral IRES Elements

The prediction of the RNA structure of the picornavirus IRES elements indicated that these IRES element could be divided into three groups (see Figs. 1–2). The PV, COX, Echo, HETV, SVDV, HRV and BEV share considerable conservation of IRES structure and these IRES elements are termed type 1 IRES. So do the EMCV, Mengo, TMEV, ERV and FMDV, their IRES elements are termed type 2. The type 1 IRES element embraces RNA structural domains A-H and J, and type 2 includes structural domains B-C and E-I. The domains D-H and J in types 1 IRES and domains E-I in the 3' portion of the IRES elements can fold to form a compact tertiary structural core by tertiary structural interactions (12), respectively. Although the two types of IRES are quite different in primary sequence, the folded structural core in the 3' portion of the IRES shares a highly similar tertiary structural feature. Among them, the domains E-H converge to form a similar three-dimensional presentation. The type 3 IRES, IRES elements of HAV, preserve the structural feature of the structural core observed in both the type 1 and type 2 IRES elements. It includes the structural domains D and J that are deleted in the type 2 IRES and the structural domain I that is deleted in the type 1 IRES. However, the RNA structure of type 3 IRES is closer to that of type 2 IRES than that of type 1 IRES.

For three types of IRES, the common feature of the structural domain C is a large multiple stem-loop structure and encompasses about 210 to 250 nt. Among them, the largest size of the domain C is observed in the type 3 IRES, and the smallest is observed in the type 1 IRES. A common structural core composed of domains D-J is just downstream of the domain C. In the common structural core, stems E and H form a pseudoknot structure and stem G is able to stack coaxially on stem F. This important structural feature is conserved in all IRES elements of picornavirus and the domains E-H are termed as the major common structural core of picornaviral IRES elements. The authentic initiator AUG is about 125 nt downstream of the structural core for PV and SVDV, 119 nt for Cox, HETV, and Echo, and 112 nt for BEV for types 1 IRES elements except that the authentic

Table 1. Sequences used in analysis

Species	Symbol	Accession No.	Species	Symbol	Accession No.
Poliovirus Type 1 (pv1)			Human Echovirus (Echo)		
Mahoney	PV1-M	J02281	Type 1/Farouk	Echo-1	X89531
Mahoney (Baltimore)	PV1-MB	V01149	Type 11	Echo-11	X80059
Sabin 1	PV1-S1	V01150	Type 12	Echo-12	X77708
P1/8-3827/Brazil/81	PV1-B8	D00260	Type 2/Cornelis	Echo-2	X77708
15/Hong Kong/81	PV1-K81	L76402	Type 22	Echo-22	S45208
NE-459/Spain/82	PV1-S82	L76409	Type 25/JV-4	Echo-25 (JV-4)	X90722
2171/USA/77	PV1-U77	D00261	Type 25	Echo-25 (M1262)	X90723
Poliovirus Type 2 (PV2)			Type 25	Echo-25 (TH222)	X90724
Sabin 2	PV2-S2	X00595	Type 3/Morrisey	Echo-3	X89533
Lansing	PV2-L2	M12197	Type 30/I	Echo-30 (I)	S76768
W-2	PV2-W-2	D00625	Type 30/P	Echo-30 (P)	S76769
364/India/56	PV2-I56	L76393	Type 4/Pesascek	Echo-4	X89534
LS2575/Kuwait/80	PV2-K80	L76403	Type 5/Noyce	Echo-5	X89535
1400/Mexico/80	PV2-M80	L76404	Type 6/Charles	Echo-6	U16283
299/USA-CA/52	PV2-U52	L76412	Type 7/2185/Finland/87	Echo-7 (2185)	L76400
Poliovirus Type 3 (PV3)			Type 7/Wallace	Echo-7 (W)	L76401
P3/Leon/37	PV3-L37	K01392	Type 8/Bryson	Echo-8	X89539
P3/Leon 12 alb	PV3-L12	X00925	Type 9/prototype Hill	Echo-9	X84981
P3/119	PV3-119	X01076	Human Enterovirus (HETV)		
P3/23127	PV3-23127	X04468	Type 71/BrCr	HETV-71 (BrCr)	U22521
P3/Sabin 3, Vaccine	PV3-S3	K00043	Type 71/MS/7423/87	HETV-MS	U22522
3054/Brazil/81	PV3-B3054	D00263	Type 70/J670/71	HETV-70	D00820
P3/119	PV3-119	X01076	Bovine Enterovirus (BEV)		
Fin/60212/85	PV3-F85	D00258	RM-2	BEV-RM-2	X79369
21267/Marocco/77	PV3-M77	L76405	ps-87	BEV-ps87	X79368
1620/Netherland/58	PV3-N58	L76406	VG-5-27	BEV-VG-5-27	D00214
17206/Netherland/70	PV3-N70	L76407	Swine Vesicular Disease Virus (SVDV)		
F29/Singapore/86	PV3-S86	L76408	H/3'76	SVDV	D00435
2142/Switzerland/80	PV3-S80	L76410	Encephalomyocarditis Virus (EMCV)		
20800/Turkey/81	PV3-T81	L76411	EMC-B	EMCV-B	M22457
50/USSR/87	PV3-U87	L76413	D/ifp-phenotype	EMCV-D2	M37588
USOL/D-BAC	PV3-USOL	D00259	EMC-D	EMCV-D2	J04335
Yunan 2/84	PV3-Y84	D00262	PV2	EMCV-PV2	X87335
Coxsackievirus Type A (Cox-A)			PV21	EMCV-PV21	X74312
A9/Griggs	Cox-A9	D00627	Ruckort	EMCV-R	M81861
A1/T.T. (Tompkins)	Cox-A1	X87584	Mengo Virus (Mengo)		
A11/Belgium 1	Cox-A11	X87592	medium plaque	Mengo	L22089
A12/Texas 12	Cox-A12	X87593	Theiler's Murine Encephalomyelitis Virus (TMEV)		
A13/Flores	Cox-A13	X87594	DA	TMEV-DA	M20301
A14/G14	Cox-A14	X87595	TO(4)	TMEV-TO4	M80885
A15/G-9	Cox-A15	X87596	TO(B15)	TMEV-TOB15	M80886
A16/G-10	Cox-A16	U05876	TO(FA)	TMEV-FA	M80883
A17/G-12	Cox-A17	X87597	TO(Yale)	TMEV-TOYale	M80890
A18/G-13	Cox-A18	X87598	BeAn 8386	TMEV-BeAn	M16020
A20/1H Pool 35	Cox-A20	X87600	GDVII	TMEV-GDV	X56019
A21/Coe	Cox-A21	D00538	MHG	TMEV-MHG	M80884
A22/Chulman	Cox-A22	X87603	Vilyuisk	TMEV-Vi	M80888
A24/EH24/70	Cox-A24	D90457	VL	TMEV-VL	M80887
A3/J.O1. (Olson)	Cox-A3	X87586	WW	TMEV-WW	M80889
A5/G.S. (Swartz)	Cox-A5	X87588	Foot and Mouth Disease Virus (FMDV)		
A7/AB-1V (Russian)	Cox-A7	X87589	Type A12/ab-119	FMDV-A12	M10975
A8/C.D. (Donovan)	Cox-A8	X87590	Type A10	FMDV-A10	M32258
Coxsackievirus Type B (Cox-B)			Asia 1	FMDV-Asia	M31578
B1	Cox-B1	M16560	Type C/serotype O1	FMDV-C-O1	M95781
B3/serotype	Cox-B3	M88483	C3R	FMDV-C3R	D10145

Table 1. (Continued)

Species	Symbol	Accession No.	Species	Symbol	Accession No.
B3/Nancy	Cox-B3 (Nancy)	M16572	C3R-O/E	FMDV-C3R-OE	D00146
B4/J.V.B.Benschoten	Cox-B4	D00149	O1-BFS	FMDV-O1-BFS	M31577
1954/B5/UK	Cox-B5	X67706	O1C	FMDV-O1C	D10143
Human Rhinovirus			O1C-O/E	FMDV-O1C-OE	D10144
Type 1B	HRV-1B	D00239	SAT1	FMDV-SAT1	M13579
Type 2	HRV-2	X02316	SAT2	FMDV-SAT2	M13580
Type 14 (Stanway)	HRV-14 (S)	K02121	SAT3	FMDV-SAT3	M13581
Type 14 (Callahan)	HRV-14 (C)	X01087	Equine Rhinovirus (ERV)		
Type 16	HRV-16	L24917	Type 1/PERV	ERV-1	X96870
Type 89	HRV-89	M16248	Type 2	ERV-2	X96871
Human Hepatitis Virus (HAV)			Human Hepatitis Virus (HAV)		
HM175/18f	HAV-18f	M59808	HM175/24A	HAV-24A	M59810
HM175/43C	HAV-43C	M59809	HM175/7MK-5	HAV-7MK	M16632
HM175/wt	HAV-wt	M14707	CF53	HAV-CF53	M63025
F.G./Italy	HAV-FG	X83302	GBM/HFS	HAV-GBM-HFS	X75216
GBM/FRhK	HAV-GBM-FRhK	X75214	GBM/wt	HAV-GBM-wt	X75215
pHAV-[1, 8, 16, 47]	HAV-pHAV	K02990	MBB/PLC/PRF/5	HAV-MBB	M20273
PA21	HAV-PA21	M63026			

initiator AUG of HRV is partially involved in the common structural core. For type 2 IRES the common structural core situates just 13–17 nt 3' to the authentic initiator AUG. The authentic initiator AUG of type 3 IRES is immediately downstream of the common structural core. This feature is like that observed in HRV IRES. The feature of the common structural core for picornavirus IRES elements is depicted in Figs. 1–2.

Phylogeny of Picornaviral IRES Elements

The unrooted phylogenetic tree of picornaviral IRES elements (Fig. 3) showed three groups as suggested by

structure comparison. The Eight sub-groups were defined by highly significant internal branches. These were coxsackievirus family, SVDV, PV, HRV, BEV, cardiovirus family, FMDV and HAV. The unclassified ERV was here assigned in the same family with EMCV, Mengo and TMEV. However, the distance between ERV and FMDV is closer than that between other members of the cardiovirus family and FMDV. The Cox-B virus IRES is closer to the IRES of Echo and HETV than that to Cox-A. The coxsackievirus family could include Cox-A, Cox-B, Echo, and HETV. The extensive sequence similarities displayed within each group suggest that intermolecular recombination and rearrangements have occurred

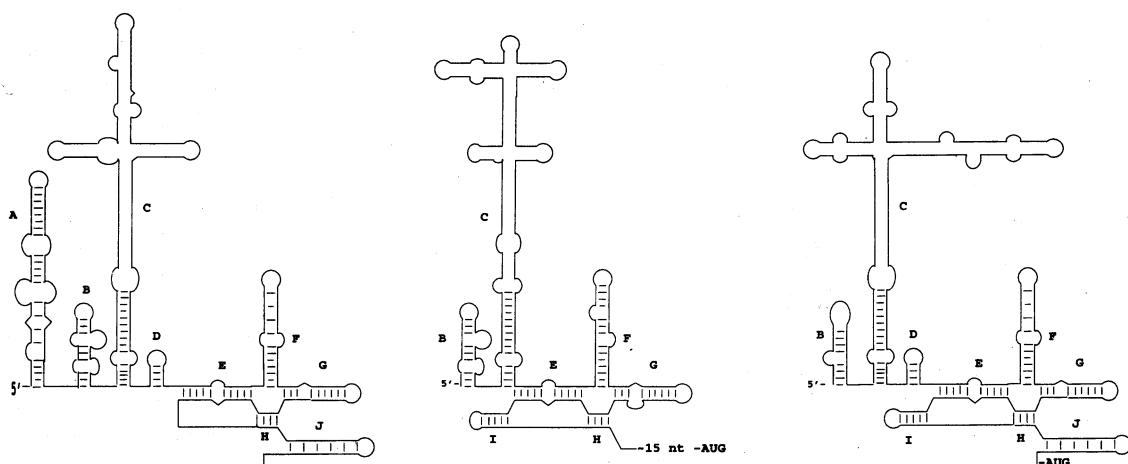


Fig. 1. Schematic diagram of common structure of picornavirus IRES. (a) Type 1 IRES, (b) Type 2 IRES and (c) Type 3 IRES.

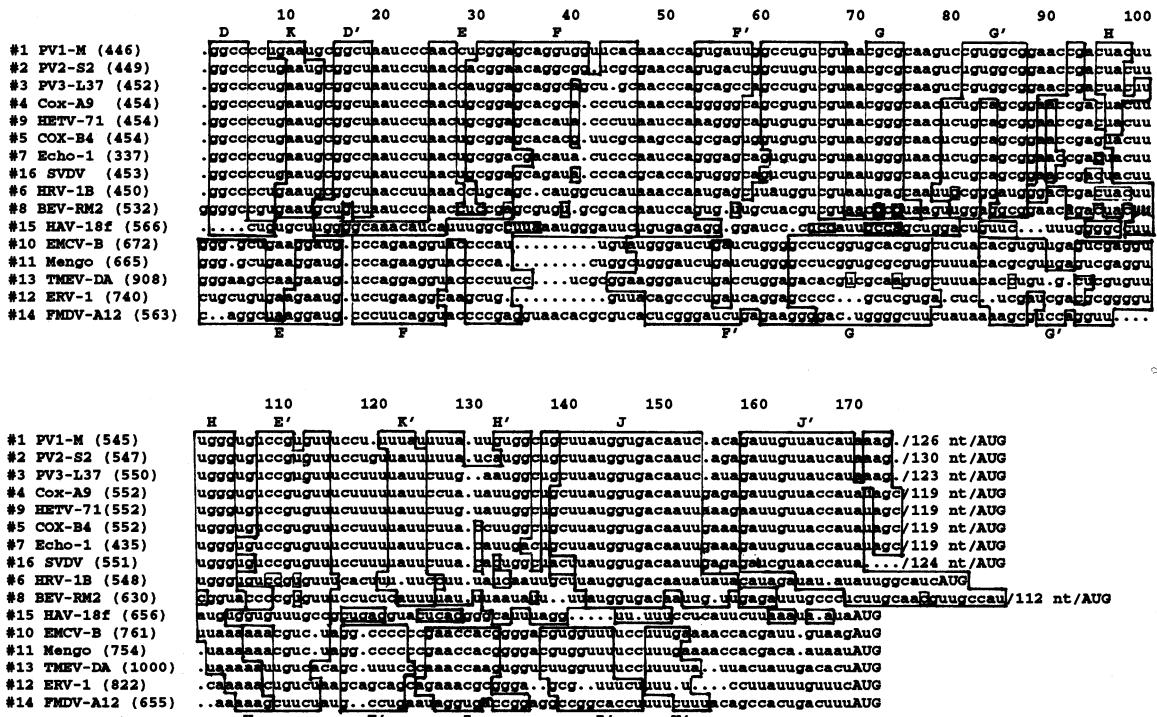


Fig. 2. Sequence alignments of the common structural core of picornavirus IRES elements. In the alignment, we selected PV1-M, PV2-S2, and PV3-L37 to represent three subtype sequences of polioviruses, respectively; Cox-A9 and Cox-B4 to represent Cox-A and Cox-B; and HRV-1B, HETV-71(BrCr), Echo-1, BEV-RM2, EMCV-B, TMEV-DA, ERV-1, FMDV-A12 and HAV-18f to represent families of HRV, HETV, Echo, BEV, EMCV, TMEV, ERV, FMDV and HAV, respectively. Deletions are denoted by dots. The conserved base pairing regions are marked by boxes and labeled by the letters D-J and D'-J'. Any bases within a box that are enclosed by another box fail to form an appropriate base pair. Additional pseudoknots predicted in types 1 IRES elements are labeled by the letters K and K'. The additional pseudoknot predicted in HAV is not labeled.

during viral IRES evolution. On the other hand, the nucleotide sequences of IRES elements are so divergent among the three types of IRES that deletions and additions must have been frequent in this genomic region during picornovirus evolution.

Possible Evolutionary Relationships of the Common Structural Core of the Picornavirus IRES

The highly structured IRES elements have been determined for all members of picornaviruses. Moreover, the major structural element in the common structural core of picornaviral IRES composed of stems E, F, G and H present not only in picornaviruses but also in HCV and pestiviruses, is consistent with their exploitation of the same internal ribosome-binding mechanism (12). The emerged common structural core should be very relevant to understanding of the molecular details of the function

of the *cis*-acting IRES element in the cap-independent translational initiation of these viruses. Phylogenetic analysis indicates that domain D in types 1 and 3 is highly conserved among members of the same group in both primary sequences and the folding patterns. Although domain D in the both types 1 and 3 is a small hairpin structure, the nt sequence of type 1 IRES is quite different from type 3. The examined sequences included 7 PV-1, 7 PV-2, 17 PV-3, 4 Cox-A, 5 Cox-B, 3 HETV, 3 BEV, 6 HRV, 18 Echo and 13 HAV viruses. The hairpin structure D in type 3 IRES is a smaller hairpin (with 3 base-pairs and 4 nt in the loop) than that of type 1. Only one base-pairs U:A was observed to be substituted by U:G in three HAV out of 13 HAV 5'UTR.

The domain J of type 3 IRES (HAV) is a small stem-loop structure and is also highly conserved in HAV. No mutation was observed in the complementary sequences of the domain J of 13 HAV. Although

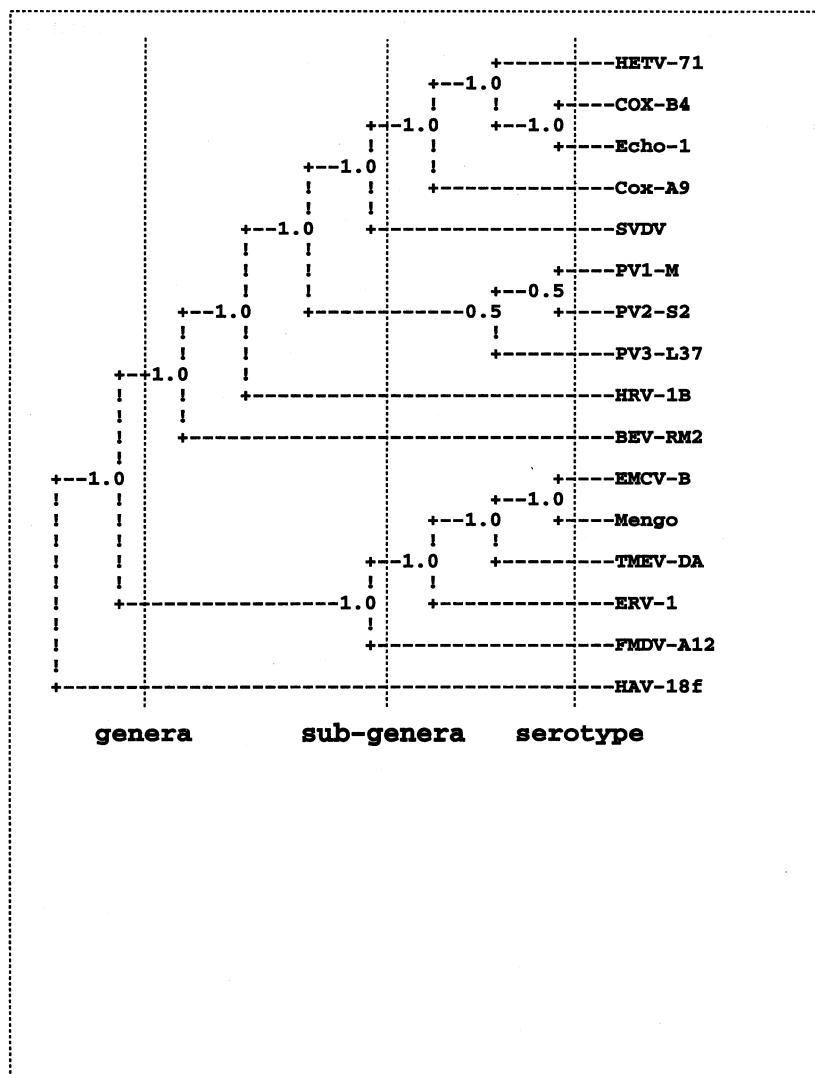


Fig. 3. Consensus phylogenetic tree obtained for the RNA sequences showed in Fig. 2. The numbers at the forks indicate the fraction of the number of times the group consisting of the sequences which are to the right of that fork occurred among four input trees. The four input trees were computed by DNA parsimony program (DNAPARS) with the jumble option to evaluate 500 random orderings of 16 input sequences.

domain J of the type 1 IRES is also a stem-loop structure, it is much greater than that of type 3 IRES. Primary sequences of domain J for type 1 IRES are highly conserved within each subgroup of type 1 IRES. The predicted folded helical stem in the domain J, indicated by the occurrence of complementary sequences (Fig. 2) are proven by two or more independent covariations occurred in the complementary sequences among these members of type 1 IRES elements. Overall, the sequence variation in the domain J of the type 1 IRES is mainly found in loops.

The structure domain I is the property of types 2 and 3 IRES. It is often presented in a hairpin structure. Numbers of independent compensatory base changes occur in the complementary sequences. In the common core structure, the domain I is located between stems E and H. It seems that the hairpin structure has a significant effect to the formation of the compact tertiary structural core. The resulting common structural core composed of domains E-I during the evolution of type 2 IRES likely improve the utilization of the type 2 IRES in the cap-independent

Tables 2–5. In the tables, the uppercase letters indicate the complementary sequences. Deletions are denoted by dashes. The identical nucleotides with the listed sequence above are indicated by dots.

Table 2. Structural conservation and possible evolutionary relationships in the element E of the structural core of picornavirus IRES elements

Table 2. (Continued).

Element E		Picornaviruses
5'-CAGGCua--AGGAuG-3'	5'-CUUCU-augCCUG-3'	FMDV-A12 (ab-119), (563--575/661--672) FMDV-C-01 (serotype O1) FMDV-O1C-OE (strain O1C-O/E) FMDV-C3R (strain C3R, FMDV-C3R) FMDV-C3R-OE (strain C3R-O/E) FMDV-A10 (type A-10) FMDV-O1-BFS (strain O1 BFS) FMDV-O1C (strain O1C)
.....	HAV-18f (HM175-18f), (576--586/659--669)
.....	HAV-wt (strain HM175 wild-type)
.....	HAV-24A (strain HM175-24A)
.....	HAV-7MK (strain HM175-7MK-5)
.....	HAV-43C (strain HM175-43C)
.....	HAV-MBB (isolate MBB)
.....	HAV-GBM-wt (strain GBM wild-type)
.....	HAV-GBM-HFS (strain GBM/HFS)
.....	HAV-GBM-FRhK (strain GBM/FRhK)
.....	HAV-pHAV (clones pHAV-[1,8,16,47])
.....	HAV-CF53
.....g.....	HAV-FG (strain F.G.)
..U.....AA.	.UU.....	HAV-pA21

Table 3. Structural conservation and possible evolutionary relationships in the element F of the structural core of picornavirus IRES elements

Element F	Picornaviruses
5'-GCAGGUGGUUCACaaaccaGUGAUUgGCCUGU-3'	PV1-M (478--509)
.....-	PV1-MB, PV1-S1
....C.-.UG...u...CA.C.....	PV1-S82
..Ga...C-.UG..g.....C..G....c....	PV1-HK
....A.-.UG....c....CA.C.a.....	PV1-B8
..G.A....C.....g.....G.Cu.UUC..	PV1-U77
A.....C....G.g.....C.....U....	PV2-S2 (481--511)
.....a...gG....u....Cg.Cca.....	PV2-L2
....Ca...gG....u....Cg..ca.....	PV2-W-2
..G.....A-C.....g.....G.CcAUUC..	PV2-M80
.....A-CUU....g....A.G..u.....	PV2-K80
....A.C...-....g.....G.....U...	PV2-U52
.....-UG....c....CA.CC.....	PV2-I56
....Ca.-CUG....c....CAGCCa.....	PV3-L37 (484--514)
.....-	PV3-119, PV3-L12, PV3-S3
.....A...-.....u.....	PV3-23127
u...A....C-CUU.....GAGG....U..u	PV3-U87
....A....U-C.U....c.....A.....U...	PV3-N70
.....A.-.UG.....CA.C.g.....	PV3-M77
.....-.....G..u.....	PV3-S80
.....-.....C.u.....	PV3-T81
.....-.....u.....	PV3-F85
....A....A...G.....a..U...	PV3-N58
....G....C-.....g.....G.u.....	PV3-S86
....C.c-.....u.....a.a.....	PV3-B86
....CA.-AUG....u....CA.C.a.....	PV3-Y84
....A....U-C.U....c....G.GA....U...	PV3-USOL
AGcaAGUGACUGCaacccaGCAGCugGCUaCU	PV3-B3054
GCACGCa-CCCUcaaaccaGGGGGcaGCGUGU	CoX-A9 (486--516)
....AU.....g.c.....U....	Cox-A16

Table 3. (Continued)

Element F	Picornaviruses
.....AU.....U.....	Cox-B1
.....AC.....g.....A.....U.....	Cox-B3, Cox-B3 (Nancy)
.....AU.....c.....U.....	Cox-B5
....A.g.UU.G....g....C.A.ug.U....	Cox-B4 (486--516)
...G.U..GUUG...u....CA.C...C...	Cox-A24
GCAAUC-GCUCACgacccaGUGAGUaGGUUGU	Cox-A21
...GAU-a.C...ca.....G.c..UC...	SVDV (485--515)
GCACAU-acCCUuaauccaAAGGGcaGUGUGU	HETV-71 (BrCr) (486--516)
.....-G..U.C..c...G....U.....	HETV-71 (MS/7423/87)
...A...-G.U.A.....GU.A.Ug..U...	HETV-70 (J670/71)
GCGUGu-GCGCACaauccaGUGuUGC-UACGU	BEV-RM-2 (565--594)
.....-.....-.....	BEV-VG-5-27, BEV-ps87
GCCAUGG-CUCAUaaaccaAUGAGcUU AUGGU	HRV-1B (482--512)
..U.GA...-A.G....c.....U.uA.C.A..	HRV-2
..c.GU...-A...C.u...G..U.uAGC....	HRV-16
...G.U...-C.....u.....G....GC...	HRV-89
...U.AU-GC..Cg.u...G..GUuG..A...	HRV-14 (Stanway and Callahan)
GCAGGU-GCUCACaauccaGUGGGUGGCCUGU	Echo-6
...A...-C.....c.....U.....	Echo-12
....A.-a.C.....-.....ca.U.....	Echo-7 (2185)
...CA...-C.CU.....AG.....a.UG...	Echo-7 (Wallace)
....C.....G....a.....A.a.....	Echo-25 (TH222)
....c-a.C.....a.....c.a.....	Echo-25 (M1262)
...CA...-C.CU.....AG.....a.UG...	Echo-9
....A.-a.C....g.a.....ca.U.....	Echo-25 (JV-4)
....A.U-..C....ug.....cA.U.....	Echo-30 (I)
--C.-U.....ua.....A..Agcc...	Echo-30 (P)
ACAUA-CUCCCaauccaGGGAGcagUGUGU	Echo-1 (371--399)
.....-C..U.....A..G.....	Echo-11
...CG...-A...g....U...ug.....	Echo-4
..G.G...-U...c.....A.G.ug.C....	Echo-8
.AG.a....A.....U.G....GC.g.	Echo-3
.GG.a..-C.A.g.g....U.G....GCC.--	Echo-2
.GGCa..-C.A.g.g....U.G....GCC.--	Echo-5
.gGAUUGGUU---.a.AACC-ugAAU.gU	Echo-22
UCCAGGAGGUacCCUUCCucgcGGAAAGGAUCUgaCCUGGA	TMEV-DA (923--964)
.....	TMEV-WW, TMEV-TO4
.....	TMEV-BeAn, TMEV-VL, TMEV-TOYale
.....u.....	TMEV-FA, TMEV-GDV
.....g.u.....	TMEV-TOB15
.....uu.....	TMEV-MHG
.....uu.....	TMEV-Vi
.....u....au.....	ERV-1 (755--789)
UCCUGAAAGGcaaGCUG--guua----CAGCCCUGaUCAGGA	ERV-2
....AC....U.c.C.GCAC (14nt) GUG.G.AU....GU....	EMCV-B (686--722)
CCCAGAAAGGUacCCCAU--ugu---AUGGGGAUCUgaUCUGGG	EMCV-D, EMCV-D2, EMCV-R
.....	EMCV-PV2, EMCV-PV21
.....c---g---c.....	Mengo (679--715)
CCCUUCAGGUacCCGAG(13nt)CUCGGGAUCUgaGAAGGG	FMDV-A12 (576--624), FMDV-O1-BFS
.....	FMDV-C-01, FMDV-O1C, FMDV-O1C-OE
.....U.....	FMDV-C3R, FMDV-C3R-OE
.....	FMDV-A10
UUUGGCCCCuuuaAUUGGGAUUCugugagaggGGAUCCCuCCAUUggcaGCUGGA	HAV-18f (587--638)
.....	HAV-24A, HAV-43C
.....	HAV-7MK, HAV-wt
.....G.....	HAV-MBB

Table 3. (Continued)

Element F	Picornaviruses
C.....ua.....	HAV-FG
C.....G.....a.....	HAV-pHAV, HAV-GBM-HFS
C.....G.....ua.....	HAV-GBM-FRhK, HAV-GBM-wt
C.....au.....a	HAV-CF53F
C.....A.G.C.....C.U.a.....aaau.....	HAV-PA21

Table 4. Structural conservation and possible evolutionary relationships in the element G of the structural core of picornavirus IRES elements

Element G	Picornaviruses
CGUaaCGCGcaagucCGUGGCG	PV1-M (510--531), PV1-MB
.....	PV1-S1, PV1-HK, PV2-U52
.....U.....	PV3-L37, PV3-119, PV3-S3, PV3-L12
.....U.....	PV1-S82, PV1-U77, PV1-B8
.....u.....	PV2-S2, PV2-K80, PV2-I56, PV3-N58
.....u.....U.....	PV2-L2, PV2-W2
.....U.....U.....	PV2-M80, PV3-B86
.....U.....U.....	PV3-M77
.....U.....	PV3-U87
.....a.....U.....	PV3-N70, PV3-23127, PV3-S80
.....a.....U.....	PV3-USOL, PV3-F85
...CU...u.u..U.....	PV3-B3054
....U....a.....U.....	PV3-T81
....G....u.....U.....	PV3-S86
....U....u.....U....AU.	PV3-Y84
CGUaaCGGGcaacUCUGcaGCG	Cox-A9 (517--538), Cox-A16
.....	Cox-B1, Cox-B3, Cox-B3 (Nancy)
.....	Cox-B4, Cox-A5, Cox-A7, Cox-A12
.....u.....	Cox-A3, Cox-A8, Cox-B5
.....U.....	Cox-A14
....AU.c....gu...AUg...	Cox-A1, Cox-A20
....A..c.A..g....Ug...	Cox-A11
....A..c.u..gu...Ug...	Cox-A22
....U.C.u..guc..UG...	Cox-A21
....C....gucC.UG...	Cox-A24
....CG....gucC.UG...	Cox-A13
....CGu..guc..UG...	Cox-A15, Cox-A17
....A..a....u.c..Ug...	Cox-A18
CGUaaUGGGcaacUCUGcaGCG	SVDV (516--537)
CGUaaUGAGcaaUUGCGggAUG	HRV-1B (513--534), HRV-2, HRV-16
.....C.u..g.G.....	HRV-89
.....C.....C.....C.	HRV-14 (stanway and Callahan)
CGUaaCgCgUAaguUGGaGGCG	BEV-RM-2 (595--616)
.....C.....	BEV-VG-5-27
.....a....cA.g....	BEV-ps87
CGUaaCGGGcaacUCUGcaGCG	HETV-71 (BrCr) (517--538)
.....U.....	HETV-MS
....U.C....guc..UG...	HETV-70 (J670/71)
CGUaaUGGGuaacUCUGcaGCG	Echo-1 (400--421), Echo-5, Echo-30 (I)
.....c.....	Echo-2, Echo-12
.....C.....	Echo-3, Echo-25 (TH222)
.....C.....C.....	Echo-4
....C....c.....	Echo-6, Ecgo-8, Echo-11
....C....c.....	Echo-7 (2185), Echo-30 (P)
.....A.c..u..C.....	Echo-7 (Wallace)

Table 4. (Continued)

Element G	Picornaviruses
.....C...c.....C.....	Echo-9, Echo-25 (JV-4)
.....g...	Echo-25 (M1262)
U...---G.Aag. uaUUC-a.UA	Echo-22
5'-GCCUCGGUGCACGUGcucuaCACGUGUugaGUCGAGGU	EMCV-B (723--760)
.....A....u...GU.....u.....	EMCV-D, EMCV-D2, EMCV-PV2
.....u....U.....u.....	EMCV-PV21
GCCUCGGUGCGCUGcuuuuaCACGCGUugaGUCGAGGU	EMCV-R
GACACGuCGCa aGUGcuuuuaCACcUGUGcuCGUGUU	Mengo (716--753)
.....G.....C.....	TMEV-DA (965--1000)
.....A...c.....U....	TMEV-ww, TMEV-T04
.....A..A..c.....U....	TMEV-TOB15
.....A..A..c.....U....	TMEV-BeAn
.....A..A..c.....U....	TMEV-VL
.....A.....u.....U....	TMEV-FA
.....A..A..u.....U....	TMEV-GDV, TMEV-TOYale, TMEV-MHG
ACACAUACACGUGCuuauaCACuUGUGcuUGUGU	TMEV-Vi
GCCCCGC-UCG-UGAcuc--UCGauCGAcGCGGGGU	ERV-1 (790--821)
.A..GuGG.GCACUG..uuuCA..UGCaGC.C..U.	ERV-2
GACUGGGGCUUcuauaaAAGCgUCCaGGUU	FMDV-A12 (625--654)
.....U.C..-A...	FMDV-Asia
.....G....u.....C..-A...	FMDV-SAT1
..U.....cc.....C..-AA..	FMDV-SAT2
.....AG.....c...CU.C..-....	FMDV-SAT3
.....A.....u.....CU.-....	FMDV-A10
.....A.....-u.....U.C..-A...	FMDV-01-BFS, FMDV-C-01
.....A.....-u.....C..-A...	FMDV-01C-OE
.....A..c..-u.a..U.C..-A...	FMDV-01C
GACUAGGGCUUcuguaaAAGUgCCU-AGUU	FMDV-C3R, FMDV-C3R-OE
GUUC-uuug-GGGC	HAV-18f (641--652), HAV-24A, HAV-wt
.....	HAV-43C, HAV-7MK, HAV-MBB, HAV-pHAV
..a.-....-g..	HAV-FG, HAV-GBM-HFS, HAV-GBM-wt
..a.-....-g..	HAV-GBM-FRhK
UUUC-uuu--GGGG	HAV-CF53, HAV-PA21

Table 5. Structural conservation and possible evolutionary relationships in the element H of the structural core of picornavirus IRES elements

Element H	Picornaviruses
5'-CGaCUACuuUGGG-3'	5'-UUUAuu-GUGGcUG-3'
.....
.....c.....	C.C....a.....
.....cUU....	...GcAA-.....
.....c.....c.a.....
.....cUU....cAA-.....
.....ac.....	...G----.....
.....c.....a.....
.....c.....C.....
.....c.....	C.....a.....
.....cUU....	C..GAA-.....
.....c.....	C..G----.....
.....c.....gua.....
.....c.....a.....
.....c.....	C.C..aua.....
.....c.....	C..G.-.a.....
5'-CGaCUACuuuGGGU-3'	5'-AUUUuuuucUGGcUG-3'
	PV1-M (536--548/568--580)
	PV1-MB, PV1-S1
	PV1-HK, PV2-23127
	PV1-S82
	PV2-S2
	PV2-L2, PV2-W-2, PV3-M77
	PV2-I56
	PV2-K80
	PV2-U52
	PV2-M80, PV3-S80,
	PV3-L37, PV3-L12, PV3-119, PV3-S3
	PV3-U87
	PV3-N70
	PV3-T81
	PV3-S86
	PV3-B86
	PV3-N58

Table 5. (Continued)

Element H	Picornaviruses
5'-AaCCGAcUAuuUGGG-3'	5'-CCUAUAUUGGcU-3'
.....a.....	.U..uUA.....
.....a.....u..	UU.-.c.....
.....a.....	.U..c.c.....
.....a.u.....	UU...uc.....
.....a.....	UU..c.....
.....a.....	UU.AuucA.....
.....a.....	.U..c.....
.....a.a.....	UU..ucA.....
.....a.....	.UC...c.....
.....a.....	.U..caA.....
.....a.....	.U....c.....
.....a.ua.UU....	UU..A.c.....
.....a.ua.UU....	UU..A.-.....
.....a.....	UU..uUA.....
.....a.....	UUC...c.....
.....a.a.....	UU..ucA.....
.....a....C..u...	.U.uGUA.....
.....c.....
AaCCGAGuacuuUGGG	.U..c.....
5'-AaCCGAcUAuuUGGG-3'	CUUAcCUUGGcU
5'-AaCCGAcUAuuUGGG-3'	5'-CUUAcAcUGGcU-3'
.....	5'-CUUGUAUUGGcU-3'
AaCCGAcUAuuUGGGA.....
5'-AGA-cUAcUU---cGGUA-3'	UUUUAuUAUGGcU
....a..a.U---g...	5'-UAUuAAUAAuuU-3'
GAACaGAcUAuuucGGUA	..C-AAC-.a...
5'-GGAcCGAcuaUUUGGG-3'	UAUUUAUCAuuUU
.....A..UA.uu....	5'-UUUAucAAUUGGcU-3'
.....A..UA.uu....UAu.....
..A..A.....u.....uUA-.....
GAcCGAcUAuuUGGGGAUUGcAUU.
5'-AacCGAcUAuuUGGG-3'	UUCA--UAUUGUC
.....	5'-CUCAcAUUGacU-3'
..C..a..A.....	..C.....G..
..C.....uA.....	..U..U.c..G..
..C.....A..u...	U...Uu...G..
..C.....u.....	U..U-U....G..
..C..a.u..U.....	U..UG--...G..
..C..a.u..U.....	.CU.u.c..G..
.....a..A..u...	..UGu.c..G..
..C.....u..U.....	U..U-U.c..G..
..C..a..A.....	..U.....G..
..C.....u..U.....	U..U..U.c..G..
..C.....u..UU....	..U.u....G..
..C.....u..U.....	U.CA.AUCu.G..
..C.....u..U.....	..UGu.....G..
CCUauCAAuCUGGU	..U.u.....
5'-CUUAUG-3'	ACUAGUUGuaAGG
.....	5'-CAuuUAGG-3'
.....
.....
.....
.....
CUUAGG	UUUAGG
	HAV-18f (653--658/687--694)
	HAV-24A, HAV-43C, HAV-wt
	HAV-7MK, HAV-MBB, HAV-FG
	HAV-pHAV, HAV-PA21
	HAV-GBM-wt, HAV-GBM-HFS
	HAV-GBM-FRhK
	HAV-CF53

Table 5. (Continued)

Element H	Picornaviruses
5'-UUAAA-3'	EMCV-B (761--765/806--810) EMCV-D, EMCV-D2, EMCV-R
....
AAAA	UUUG
AAAA	UUUG
5'-AAAA-3'	Mengo (754--757/798--801) ERV-1 (824--827/864--867)
....G
5'-AAAG-3'	ERV-2
....	FMDV-A12 (657--660/701-704)
....	FMDV-C3R, FMDV-C3R-OE
.a..	FMDV-SAT1, FMDV-SAT2
...c.	FMDV-A10
...c	FMDV-SAT-3
A....	FMDV-O1, FMDV-C-O1
A....	FMDV-O1C, FMDV-O1C-OE
AAAAG	FMDV-O1-BFS, FMDV-Asia
5'-AAAAAA-3'	TMEV-DA (1001--1006/1046--1051) TMEV-BeAn, TMEV-VL, TMEV-GDV
.....	TMEV-MHG, TMEV-FA
.....	TMEV-TOB15, TMEV-TOYale
....a..	TMEV-WW, TMEV-TO4
UUAA	TMEV-Vi

internal translation for EMCV, Mengo, TMEV, ERV and FMDV.

The variation and substitution in the structural domains E, F, G, and H are summarized in Tables 2–5, respectively. Structural domains F and G are often a hairpin structure. The domains E and H fold together to form a pseudoknot. Each structural element is highly conserved among the IRES elements of the same group but are quite different in primary sequence and in the size of both the stem and loop regions between different groups. The substantial sequence-to-sequence differences in length for these three types of IRES occur in the hairpin structures of the common structural core. In the most extreme case of hairpin F, the extra nucleotides form hairpin structures for which there are no homologs with the shorter sequences. Typically, hairpin F of type 3 IRES includes 20 more nt than the hairpin in the type 1 IRES. For the case of HAV-18f IRES, the extra nucleotides extend both the helical stems and loops. Even if hairpin F is compared within the same group, it is the most variable element in the core structure. For instance, the hairpin F (16 base-pairs) of ERV-2 IRES contains 15, 13, and 8 more nt than that of ERV-1, EMCV-B and Mengo, and TMEV-DA, respectively in which the lengths of hairpin F are 13, 15, and 17 base-pairs, respectively. Numerous independent covariations occur in the complementary sequences of

these hairpin F such that their pairing potential is preserved, and the extra nt are mainly absorbed by the big hairpin loop of ERV-2. This high variation in the nt sequence but, nevertheless, high conservation of RNA folding in the structural domain F is one of the most compelling arguments for the crucial role of these structures in the cap-independent translational initiation of picornaviruses.

All in all, most point mutations found in the complementary sequences of domains E-H in the each IRES group either did not perturb the structures significantly or were accompanied by compensatory mutations. This suggests that variant IRES selected in vivo evolved by covariant substitution in the complementary sequences, and by the gradual addition and deletion of some structural domains and/or elements to facilitate the *cis*-acting function of IRES in the specific host-cell environment.

The main common structural motif of picornavirus IRES, composed of domains E, F, G, and H is also preserved in the IRES elements of HCV and pestiviruses (12). However, the IRES elements of HCV and pestiviruses contain ~320 nt and are much shorter than picornavirus IRES. The IRES sequences also diverge significantly from picornavirus IRES. Furthermore, the IRES sequences of HCV and pestiviruses have been reported including ~30 nt coding sequence (17). This feature is so different from

the picornavirus IRES that the IRES elements of HCV and pestiviruses should be classified as an independent group. At present, one cannot determine the exact rooting of the evolution tree of picornavirus IRES elements. Further studies will provide insights into the origin and evolution of viral IRES elements.

Acknowledgments

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

1. Pelletier J. and Sonenberg N., *Nature* 334, 320–325, 1988.
2. Nicholson R., Pelletier J., Le S.-Y., and Sonenberg N., *J Virol* 65, 5886–5894, 1991.
3. Jang S.K., Krausslich H.-G., Nicklin M.J.H., Duke G.M., Palmenberg A.C., and Wimmer E., *J Virol* 62, 2636–2643, 1988.
4. Jang S.K., Davies M.V., Kaufman R.J., and Wimmer E., *J Virol* 63, 1651–1660, 1989.
5. Bandyopadhyay P.K., Wang C., and Lipton H.L., *J Virol* 66, 6249–6256, 1992.
6. Borman A. and Jackson R.J., *Virology* 188, 685–696, 1992.
7. Kuhn R., Luz N., and Beck E., *J Virol* 64, 4625–4631, 1990.
8. Brown E.A., Day S.P., Jansen R.W., and Lemon S.M., *J Virol* 65, 5828–5838, 1991.
9. Glass M.J. and Summers D.F., *Virology* 193, 1047–1050, 1993.
10. Wimmer E., Hellen C.U.T., and Cao X., *Annu Rev Genet* 27, 353–436, 1993.
11. Ehrenfeld E. and Semler B., *Curr Top Microbiol Immunol* 203, 65–83, 1995.
12. Le S.-Y., Siddiqui A., and Maizel Jr. J.V., *Virus Genes* 12, 135–147, 1996.
13. Wilbur W.J. and Lipman D.J., *Proc Natl Acad Sci USA* 80, 726–730, 1983.
14. Le S.-Y. and Zuker M., *J Mol Biol* 216, 729–741, 1990.
15. Le S.-Y., Chen J.-H., and Maizel Jr. J.V., *Nucl Acids Res* 21, 2173–2178, 1993.
16. Felsenstein J., *Annu Rev Genet* 22, 521–565, 1988.
17. Reynolds J.E., Kaminski A., Kettinen H.J., Grace K., Clarke B.E., Carroll A.R., Rowlands D.J., and Jackson R.J., *EMBO J* 14, 6010–6020, 1995.